1. Attempt *any three* of the following:

a. Define GIS. Briefly explain any two capabilities of GIS. (Reference Book Page 32)

A GIS is a computer-based system that provides the following four sets of capabilities to handle georeferenced data:

1. *Data capture and preparation*
2. *Data management*, including storage and maintenance
3. *Data manipulation and analysis*
4. *Data presentation*

b. What is GI System, GI Science and GIS Application? Explain. (Reference book Page 43)

The discipline that deals with all aspects of the handling of spatial data and Geo-information is called *geographic information science* (often abbreviated to geo-information science or just GIScience).

Geo-Information Science is the scientific field that attempts to integrate different disciplines studying the methods and techniques of handling spatial information.

The difference between a geographic information system and and a GIS *application*. The same software package that we used to do this analysis could also be used to analyse forest plots in northern Thailand, for instance. That would be a different application, but would make use of the same software. GIS software can (generically) be applied to many different applications. When there is no risk of ambiguity, people sometimes do not make the distinction between a 'GIS' and a 'GIS application'.

c. How modeling helps in representing real world? Explain. (Reference book Page 49)

'Modelling' is a term used in many different ways and which has many different meanings. A representation of some part of the real world can be considered a *model* because the representation will have certain characteristics in common with the real world. Specifically, those which we have identified in our model design. This then allows us to study and operate on the model itself instead of the real world in order to test what happens under various conditions, and help us answer 'what if' questions.

Models—as representations—come in many different flavours. In the GIS environment, the most familiar model is that of a *map*. A map is a miniature representation of some part of the real world. Paper maps are the most common, but digital maps also exist. Databases are another important class of models. A database can store a considerable amount of data, and also provides various functions to operate on the stored data. The collection of stored data represents some real world phenomena, so it too is a model. Obviously, here we are especially interested in databases that store spatial data. Digital models (as in a database or GIS) have enormous advantages over paper models (such as maps). They are more flexible, and therefore more easily changed for the purpose at hand. In principle, they allow animations and simulations to be carried out by the computer system. This has opened up an important toolbox that can help to improve our understanding of the world.

Most maps and databases can be considered *static models*. At any point in time, they represent a single state of affairs. Usually, developments or changes in the real world are not easily recognized in these models. *Dynamic models* or -*process models* address precisely this issue. They emphasize changes that have taken place, are taking place or may take place sometime in the future. Dynamic models are inherently more complicated than static models, and usually require much more computation. Simulation models are an important class of dynamic models that allow the simulation of real world processes.

E-next THE NEXT LEVEL OF EDUCATION

d. Define Geographic field. Explain different data types and values. (Reference book Page 72 & 75)

A field is a geographic phenomenon that has a value 'everywhere' in the study area.

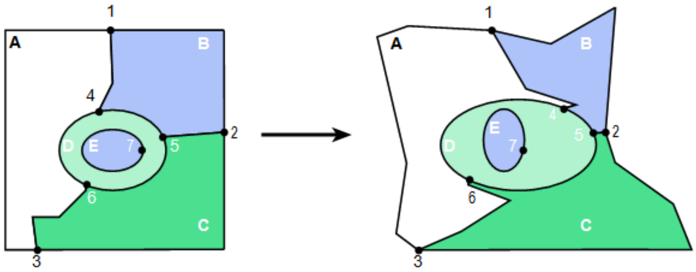
Data types and values

1. *Nominal data values*
2. *Ordinal data values*
3. *Interval data values*
4. *Ratio data values*

e. **Write a note on Topology and spatial relationships. (Reference book Page 101)**

Topology deals with spatial properties that do not change under certain transformations.

- Area *E* is still inside area *D*,
- The neighbourhood relationships between *A*, *B*, *C*, *D*, and *E* stay intact, and their boundaries have the same start and end nodes, and
- The areas are still bounded by the same boundaries, only the shapes and lengths of their perimeters have changed.



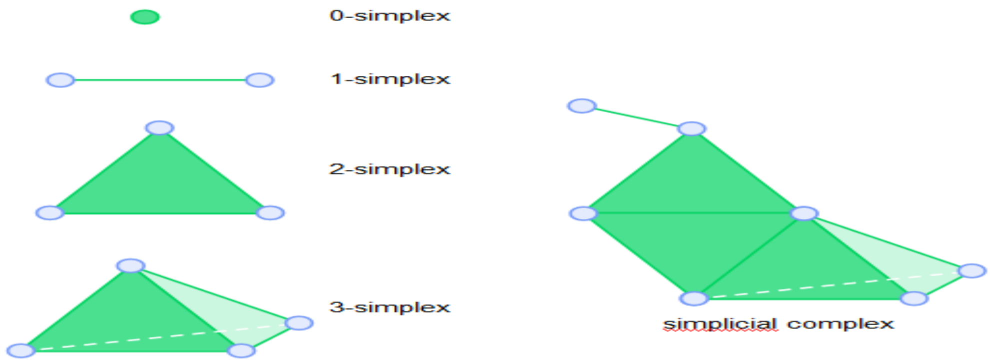
Topology refers to the spatial relationships between geographical elements in a data set that do not change under a continuous transformation. Topological relationships are built from simple elements into more complex elements: nodes define line segments, and line segments connect to define lines, which in turn define polygons. The fundamental issues relating to order, connectivity and adjacency of geographical elements form the basis of more sophis- ticated GIS analyses. These relationships (called topological properties) are invariant under a continuous transformation, referred to as a topological mapping.

Topological relationships

The mathematical properties of the geometric space used for spatial data can be described as follows:

E-next
THE NEXT LEVEL OF EDUCATION

- The space is a three-dimensional *Euclidean space* where for every point we can determine its three-dimensional coordinates as a triple (x, y, z) of real numbers. In this space, we can define features like points, lines, polygons, and volumes as geometric primitives of the respective dimension. A point is zero-dimensional, a line one-dimensional, a polygon two-dimensional, and a volume is a three-dimensional primitive.
- The space is a *metric space*, which means that we can always compute the distance between two points according to a given distance function. Such a function is also known as a *metric*.
- The space is a *topological space*, of which the definition is a bit complicated. In essence, for every point in the space we can find a neighbourhood around it that fully belongs to that space as well.
- *Interior* and *boundary* are properties of spatial features that remain invariant under topological mappings. This means, that under any topological mapping, the interior and the boundary of a feature remains unbroken and intact.



f. **Explain the temporal dimension using suitable example. (Reference book Page 126)**

Geographic phenomena are also *dynamic*; they change over time.

Examples of the kinds of questions involving time include:

- Where and when did something happen?
- How fast did this change occur?
- In which order did the changes happen?

Spatiotemporal data models are ways of organizing representations of space *and* time in a GIS.

Different ‘concepts’ of time.

- **Discrete and continuous time**:
- **Valid time and transaction time**:
- **Linear, branching and cyclic time**:
- **Time granularity**
- **Absolute and relative time**:

2. Attempt *<u>any three</u>* of the following:

a. List the functional components of GIS. Explain any two of them in details. (Reference book Page 144)

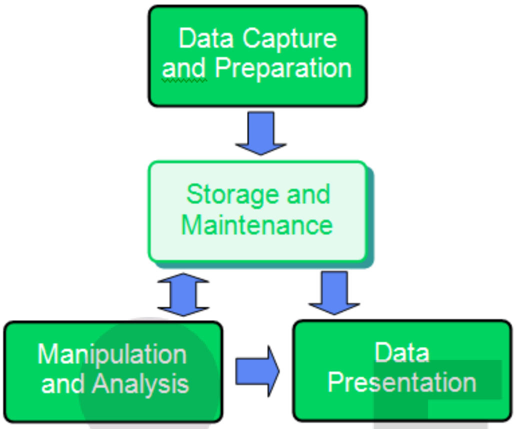


Figure 3.1: Functional components of a GIS

I-next
THE NEXT LEVEL OF EDUCATION

b. Explain the various reasons for using DBMS in GIS. (Reference book Page 158)

A *database* is a large, computerized collection of structured data.

A *database management system* (DBMS) is a software package that allows the user to set up, use and maintain a database.

There are various reasons why one would want to use a DBMS for data storage and processing.

- A DBMS supports the storage and manipulation of *very large data sets*.

Some data sets are so big that storing them in text files or spreadsheet files becomes too awkward for use in practice. The result may be that finding simple facts takes minutes, and performing simple calculations perhaps even hours. A DBMS is specifically designed for this purpose.

- A DBMS can be instructed to guard over *data correctness*.

For instance, an important aspect of data correctness is data entry checking: ensuring that the data that is entered into the database does not contain obvious errors. For instance, since we know the study area we are working in, we also know the range of possible geographic coordinates, so we can ensure the DBMS checks them.This is a simple example of the type of rules, generally known as *integrity constraints* that can be defined in and automatically checked by a DBMS. More complex integrity constraints are certainly possible, and their definition is part of the design of a database.

- A DBMS supports the *concurrent use* of the same data set by many users
- A DBMS provides a high-level, *declarative query language*.
- A DBMS supports the use of a *data model*. A data model is a language with which one can define a database structure and manipulate the data stored in it.
- A DBMS includes *data backup* and *recovery* functions to ensure data availability at all times.
- A DBMS allows the control of *data redundancy*. A well-designed database takes care of storing single facts only once. Storing a fact multiple times—a phenomenon known as *data redundancy*—can lead to situations in which stored facts may contradict each other, causing reduced usefulness of the data.

c. **Write a note on Spatial Data functionality. (Reference book Page 182)**

A *spatial database* allows users to store, query and manipulate collections of spatial data.

DBMS vendors have over the last 20 years recognized the need for storing more complex data, like spatial data. The main problem was that there is additional functionality needed by DBMS in order to process and manage spatial data. As the capabilities of our hardware to process information has increased, so too has the desire for better ways to represent and manage spatial data. During the 1990's, *object-oriented* and *object-relational* data models were developed for just this purpose. These extend standard relational models with support for objects, including 'spatial' objects.

Currently, GIS software packages are able to store spatial data using a range of commercial and open source DBMSs such as Oracle, Informix, IBM DB2, Sybase, and PostgreSQL, with the help of spatial extensions. Some GIS software have integrated database 'engines', and therefore do not need these extensions. ESRI's ArcGIS, for example, has the main components of the MS Access database software built-in. This means that the designer of a GIS application can choose whether to store the application data in the GIS or in the DBMS. Spatial databases, also known as *geodatabases*,[3] are implemented directly on existing DBMSs, using extension software to allow them to handle spatial objects spatial data can be stored in a special database column, known as the geometry column, (or feature or shape, depending on the specific software package),as shown below.

Parcel	PId	Geometry	OwnerID
	3421	"MULTIPOLYGON(((257462.704979333 464780.750851061,257463.89798...)))"	435
	8871	"MULTIPOLYGON(((257409.813950544 464789.91585049,257407.896903...)))"	550
	2109	"MULTIPOLYGON(((257785.714911912 464796.839972167,257782.59794...)))"	1040
	1515	"MULTIPOLYGON(((257790.672100448 464807.13792585,257788.608078...)))"	245
	3434	"MULTIPOLYGON(((257435.527950478 464803.92887633,257428.254887...)))"	486
	6371	MULTIPOLYGON(((257432.476077854 464813.848852072,257433.147910...)))"	950
	2209	"MULTIPOLYGON(((257444.888027332 464826.555046319,257446.43201...)))"	1840
	1505	"MULTIPOLYGON(((256293.760107491 464935.203846095,256292. 00881...)))"	145

This means GISs can rely fully on DBMS support for spatial data, making use of a DBMS for data query and storage (and multi-user support), and GIS for spatial functionality. Small-scale GIS applications may not require a multi-user capability, and can be supported by spatial data support from a personal database.

A geodatabase allows a wide variety of users to access large data sets (both geographic and alphanumeric), and the management of their relations, guaranteeing their integrity. The Open Geospatial Consortium (OGC) has released a series of standards relating to geodatabases that (amongst other things), define:

- Which tables must be present in a spatial database (i.e. geometry columns
- table and spatial reference system table)
- The data formats, called 'Simple Features' (i.e. point, line, polygon, etc.)
- A set of SQL-like instructions for geographic analysis.

The architecture of a spatial database differs from a standard RDBMS not only because it can handle geometry data and manage projections, but also for a larger set of commands that extend standard SQL language (e.g. distance calculations, buffers, overlay, conversion between coordinate systems, etc.).

d. **Explain the relational data model using suitable example. (Reference book Page 164)**

A *data model* is a language that allows the definition of:

- The *structures* that will be used to store the base data,
- The *integrity constraints* that the stored data has to obey at all moments in time, and
- The *computer programs* used to manipulate the data.

For the *relational data model*, the structures used to define the database are *attributes*, *tuples* and *relations*.

Relations, tuples and attributes

In the relational data model, a database is viewed as a collection of *relations*, also known as *tables*.

A table or relation is itself a collection of *tuples* (or records). In fact, each table is a

THE NEXT LEVEL OF EDUCATION

	collection of tuples *that are similarly shaped*. By this, we mean that a tuple has a fixed number of named fields, also known as attributes. All tuples in the same relation have the same named fields. In a diagram, as in Figure below, relations can be displayed as tabular form data. An *attribute* is a named field of a tuple, with which each tuple associates a value, the tuple's *attribute value*. The example relations provided in the figure should clarify this. The Private-Person table has three tuples; the Surname attribute value for the first tuple illustrated is 'Garcia.' The phrase 'that are similarly shaped' takes this a little bit further. It requires that all values for the same attribute come from a single domain of values. An attribute's *domain* is a (possibly infinite) set of atomic values such as the set of integer number values, the set of real number values, etc. In our example cadastral database, the domain of the Surname attribute, for instance, is string, so any surname is represented as a sequence of text characters, i.e. as a string. The availability of other domains depends on the DBMS, but usually integer (the whole numbers), real (all numbers), date, yes/no and a few more are included.

PrivatePerson	TaxId	Surname	BirthDate
	101-367	Garcia	10/05/1952
	134-788	Chen	26/01/1964
	101-490	Fakolo	14/09/1931

Parcel	PId	Location	AreaSize
	3421	2001	435
	8871	1462	550
	2109	2323	1040
	1515	2003	245

TitleDeed	Plot	Owner	DeedDate
	2109	101-367	18/12/1996
	8871	101-490	10/01/1984
	1515	134-788	01/09/1991
	3421	101-367	25/09/1996

e. **Differentiate between Vector data and Raster Data. (Reference book Page 152)**

Raster representation	*Vector representation*
Advantages	
• Simple data structure • Simple implementation of overlays • Efficient for image processing	• efficient representation of topology • adapts well to scale changes • allows representing networks • allows easy association with attribute data
Disadvantages	
• Less compact data structure • Difficulties in representing Topology • Cell boundaries independent of feature boundaries	• complex data structure • overlay more difficult to implement • inefficient for image processing • more update-intensive

f. **Write a note on Spatial Data Infrastructure. (Reference book Page 146)**

For reasons that include efficiency and legislation, many organizations are forced to work in a cooperative setting in which geographic information is obtained from, and provided to, partner organizations and the general public. The sharing of spatial data between the various GISs in those organizations is of key importance and aspects of data dissemination, security, copyright and pricing require special attention. The design and maintenance of a Spatial Data Infrastructure (SDI) deals with these issues.

An SDI is defined as "the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data". Fundamental to those arrangements are—in a wider sense—the agreements between organizations and in the narrow sense, the agreements between software systems on *how* to share the geographic information. In SDI, standards are often the starting point for those agreements. Standards exist for

E-next THE NEXT LEVEL OF EDUCATION

all facets of GIS, ranging from data capture to data presentation. They are developed by different organizations, of which the most prominent are the International Organization for Standardisation (ISO) and the Open Geospatial Consortium (OGC).
Typically, an SDI provides its users with different facilities for finding, viewing, downloading and processing data. Because the organizations in an SDI are normally widely distributed over space, computer networks are used as the means of communication. With the development of the internet, the functional components of GIS have been gradually become available as web-based applications. Much of the functionality is provided by so called geo-webservices, software programs that act as an intermediate between geographic data(bases) and the users of the web. Geo-webservices can vary from a simple map display service to a service which involves complex spatial calculations. For their spatial data handling, these services commonly use standardized raster and vector representations following the above mentioned standards.

3. Attempt _any three_ of the following:

a. What are the different classifications of Map Projections? Explain any two. (Reference book Page 220)

A *map projection* is a mathematically described technique of how to represent the Earth's curved surface on a flat map.

Classification of map projections

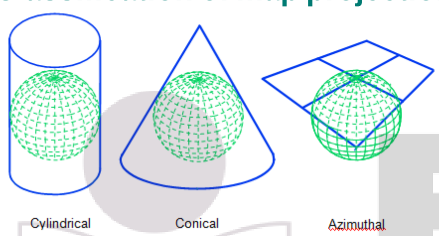


Figure 4.16: Classes of map projections

E-next
THE NEXT LEVEL OF EDUCATION

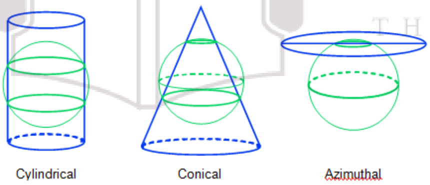


Figure 4.17: Three secant projection classes

b. Write a note on GPS. (Reference book Page 43)

The NAVSTAR Global Positioning System (GPS) was declared operational in 1994, providing Precise Positioning Services (PPS) to US and allied military forces as well as US government agencies, and Standard Positioning Services (SPS) to civilians throughout the world. Its space segment nominally consists of 24 satellites, each of which orbit our planet in 11h58m at an altitude of 20,200 km. There can be any number of satellites active, typically between 21 and 27. The satellites are organized in six orbital planes, somewhat irregularly spaced, with an angle of inclination of 55–63° with the equatorial plane, nominally having four satellites each (see Figure 4.28). This means that a receiver on Earth will have between five and eight (sometimes up to twelve) satellites in view at any point in time. Software packages exist to help in planning GPS surveys, identifying expected satellite set-up for any location and time.
GPS's control segment has its master control in Colorado, US, and monitor stations in a belt around the equator, namely in Hawaii, Kwajalein Atoll in the Marshall Islands, Diego Garcia (British Indian Ocean Territory) and Ascension Island (UK, southern Atlantic Ocean).
The NAVSTAR satellites transmit two radio signals, namely the L1 frequency at 1575.42 MHz and the L2 frequency at 1227.60 MHz. There are also a third and fourth signal, but they are not important for our discussion here. The first two signals consist of:

- The carrier waves at the given frequencies,
- A coarse ranging code, known as C/A, modulated on L1,

- An encrypted precision ranging code, known as P(Y), modulated on L1 and L2, and
- A navigation message modulated on both L1 and L2.

The role of L2 is to provide a second radio signal, thereby allowing (the more expensive) dual-frequency receivers a way of determining fairly precisely the actual ionospheric delay on satellite signals received. The role of the ranging codes is two-fold:

1. To identify the satellite that sent the signal, as each satellite sends unique codes, and the receiver has a look-up table for these codes, and
2. To determine the signal transit time, and thus the satellite's pseudorange.

The navigation message contains the satellite orbit and satellite clock error information, as well as some general system information. GPS also carries a fifth, encrypted military signal carrying the M-code. GPS uses WGS84 as its reference system. It has been refined on several occasions and is now aligned with the ITRF at the level of a few centimetres worldwide. GPS has adopted UTC as its time system.

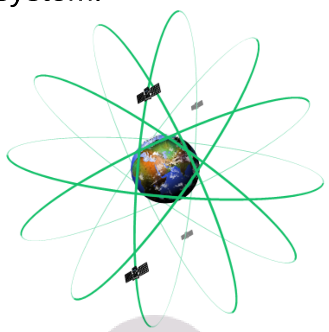
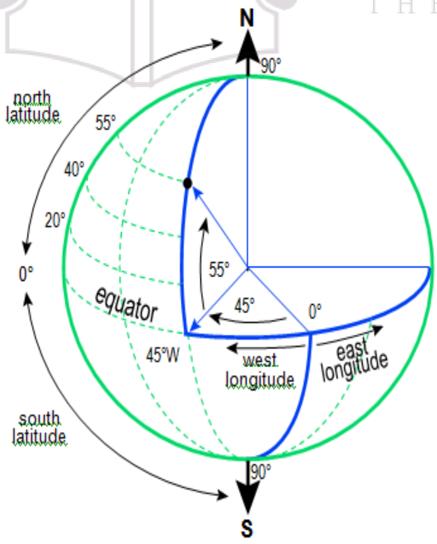


Figure 4.28: Constellation of satellites, four shown in only one orbit plane, in the GPS system.

c. **Explain 2D geographic coordinate system using suitable example. (Reference book Page 207)**

2D Geographic coordinates (φ, λ)

The most widely used global coordinate system consists of lines of geographic *latitude* (phi or φ or ϕ) and *longitude* (lambda or λ). Lines of equal latitude are called parallels. They form circles on the surface of the ellipsoid. Lines of equal longitude are called meridians and they form ellipses (meridian ellipses) on the ellipsoid.



The latitude (φ) of a point P is the angle between the ellipsoidal normal through P' and the equatorial plane. Latitude is zero on the equator ($\varphi = 0^\circ$), and increases towards the two poles to maximum values of $\varphi = +90^\circ$ (N 90°) at the North Pole and $\varphi = -90^\circ$ (S 90°) at the South Pole.

The longitude (λ) is the angle between the meridian ellipse which passes through Greenwich and the meridian ellipse containing the point in question. It is measured in the equatorial plane from the meridian of Greenwich ($\lambda = 0^\circ$) either eastwards through $\lambda = +180^\circ$ (E 180°) or westwards through $\lambda = -180^\circ$ (W 180°).

Latitude and longitude represent the geographic coordinates (φ, λ) of a point P' with respect to the selected reference surface. They are always given in angular units. For example, the coordinates for City hall in Enschede are $\varphi = 52^\circ 13\ 26.2\ N,\ \lambda = 6^\circ 53\ 32.1\ E$

The graticule on a map represents the projected position of the geographic coordinates (φ, λ) at constant intervals, or in other words the projected position of selected meridians and parallels. The shape of the graticule depends largely on the characteristics of the map projection and the scale of the map.

d. **What is trend surface fitting? Explain. (Reference book Page 326)**

In trend surface fitting, the assumption is that the entire study area can be represented by a formula $f(x, y)$ that for a given location with coordinates (x, y) will give us the approximated value of the field in that location.

The key objective in trend surface fitting is to derive a formula that best describes the field. Various classes of formulæ exist, with the simplest being the one that describes a flat, but tilted plane:

$$f(x, y) = c_1 \cdot x + c_2 \cdot y + c_3.$$

If we believe—and this judgement must be based on domain expertise—that the field under consideration can be best approximated by a tilted plane, then the problem of finding the best plane is the problem of determining best values for the coefficients c_1, c_2 and c_3. This is where the point measurements earlier obtained become important. Statistical techniques known as *regression techniques* can be used to determine values for these coefficients c_i that best fit with the measurements. A plane will be fitted through the measurements that makes the smallest overall error with respect to the original measurements. In Figure 5.15, we have used the same set of point measurements, with four different approximation functions. Part (a) has been determined under the assumption that the field can be approximated by a tilted plane, in this case with a downward slope to the southeast. The values found by regression techniques were: $c_1 = -1.83934$, $c_2 = 1.61645$ and $c_3 = 70.8782$, giving us:

$$f(x, y) = -1.83934 \cdot x + 1.61645 \cdot y + 70.8782.$$

Clearly, not all fields are representable as simple, tilted planes. Sometimes, the theory of the application domain will dictate that the best approximation of the field is a more complicated, higher-order polynomial function. Three such functions were the basis for the fields illustrated in Figure 5.15(b)–(d).

The simplest extension from a tilted plane, that of *bilinear saddle*, expresses some dependency between the x and y dimensions:

E-next
THE NEXT LEVEL OF EDUCATION

$$f(x, y) = c_1 \cdot x + c_2 \cdot y + c_3 \cdot xy + c_4.$$

This is illustrated in part (b). A further step up the ladder of complexity is to consider *quadratic surfaces*, described by:

$$f(x, y) = c_1 \cdot x^2 + c_2 \cdot x + c_3 \cdot y^2 + c_4 \cdot y + c_5 \cdot xy + c_6.$$

The objective is to find six values for our coefficients that best match with the measurements. A bilinear saddle and a quadratic surface have been fitted through our measurements in Figure 5.15(b) and (c), respectively.

Part (d) of the figure illustrates the most complex formula of the surfaces in Figure 5.15, the *cubic surface*. It is characterized by the following formula:

$$\begin{aligned} f(x, y) = {} & c_1 \cdot x^3 + c_2 \cdot x^2 + c_3 \cdot x + \\ & c_4 \cdot y^3 + c_5 \cdot y^2 + c_6 \cdot y + \\ & c_7 \cdot x^2 y + c_8 \cdot xy^2 + c_9 \cdot xy + c_{10}. \end{aligned}$$

Trend surface fitting is a useful technique of continuous field approximation, though determining the 'best fit' values for the coefficients c_i is a time-consuming operation, especially with many point measurements. Once these best values have been determined, we know the formula, making it possible to compute an approximated value for any location in the study area.

It is possible to use trend surfaces for both *global* and *local* trends. Global trend surface fitting is based on the assumption that the entire study area can be approximated by the same mathematical surface. However in many cases, the assumption that a single formula can describe the field for the *entire* study area is an unrealistic one. Capturing all the fluctuation of a natural geographic field in a reasonably sized study area, demands polynomials of extreme orders, and these quickly become computationally impossible to decipher.

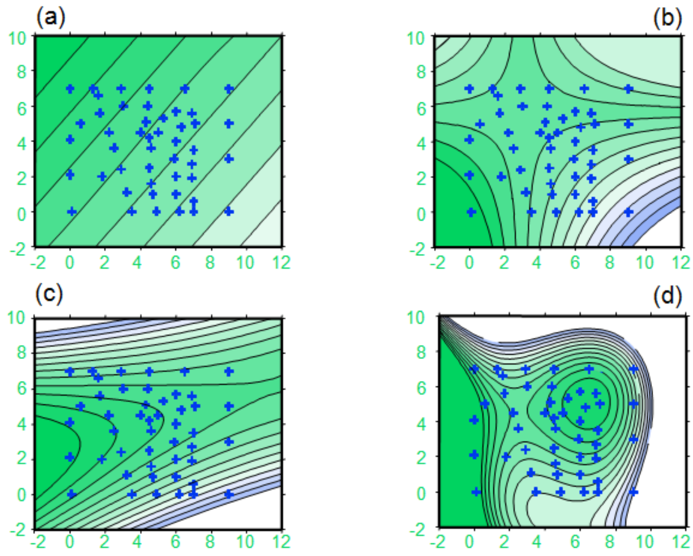


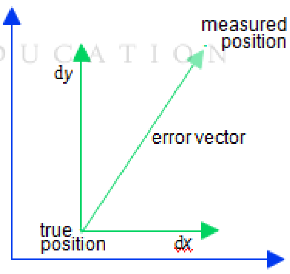
Figure 5.15: Various global trend surfaces obtained from regression techniques: (a) simple tilted plane; (b) bilinear saddle; (c) quadratic surface; (d) cubic surface. Values range from white (low), via black and light grey to dark grey (high).

e. **How Root Mean Square is used to mean location accuracy? Explain. (Reference book Page 289)**

Root mean square error

Location accuracy is normally measured as a *root mean square error* (*RMSE*). The RMSE is similar to, but not to be confused with, the standard deviation of a statistical sample. The value of the RMSE is normally calculated from a set of check measurements (coordinate values from an independent source of higher accuracy for identical points). The differences at each point can be plotted as error vectors, as is done in Figure 5.3 for a single measurement. The error vector can be seen as having constituents in the *x*- and *y*-directions, which can be recombined by vector addition to give the error vector representing its locational error.

For each checkpoint, the error vector has components δx and δy. The observed errors should be checked for a *systematic* error component, which may indicate a (possibly repairable) lapse in the measurement method. Systematic error has occurred when $\delta x = 0$ or $\delta y = 0$.



The systematic error δx in x is then defined as the average deviation from the true value:

$$\delta x = \frac{1}{n} \sum_{i=1}^{n} \delta x_i.$$

Analogously to the calculation of the variance and standard deviation of a statistical sample, the root mean square errors m_x and m_y of a series of coordinate measurements are calculated as the square root of the average squared deviations:

$$m_x = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \delta x_i^2} \quad \text{and} \quad m_y = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \delta y_i^2},$$

where δx^2 stands for $\delta x \cdot \delta x$. The total RMSE is obtained with the formula

$$m_{\text{total}} = \sqrt{m_x^2 + m_y^2},$$

which, by the Pythagorean rule, is the length of the average (root squared) vector.

f. **Write a note on Krigging. (Reference book Page 43)**

Kriging was originally developed my mining geologists attempting to derive accurate estimates of mineral deposits in a given area from limited sample measurements. It is an advanced interpolation technique belonging to the field of *geostatistics* , which can deliver good results if applied properly and with enough sample points. Kriging is usually used when the variation of an attribute and/or the density of sample points is such that simple methods of interpolation may give unreliable predictions.

Kriging is based on the notion that the spatial change of a variable can be described as a function of the distance between points. It is similar to IDW interpolation, in that it the surrounding values are weighted to derive a value for an unmeasured location. However, the kriging method also looks at the overall spatial arrangement of the measured points and the spatial correlation between their values, to derive values for an unmeasured location.

The first step in the kriging procedure is to compare successive pairs of point measurements to generate a *semi-variogram*. In the second step, the semi-variogram is used to calculate the weights used in interpolation. Although kriging is a powerful technique, it should not be applied without a good understanding of geostatistics, including the principle of spatial autocorrelation.

4. Attempt *<u>any three</u>* of the following:

a **List the four classifications of analytical functions of GIS. Explain any one in details. (Reference book Page 344)**

There are many ways to classify the analytical functions of a GIS.

1. **Classification, retrieval, and measurement functions.** All functions in this category are performed on a single (vector or raster) data layer, often using the associated attribute data.
    - Classification allows the assignment of features to a class on the basis of attribute values or attribute ranges (definition of data patterns). On the basis of reflectance characteristics found in a raster, pixels may be classified as representing different crops, such as potato and maize.
    - Retrieval functions allow the selective search of data. We might thus retrieve all agricultural fields where potato is grown.
    - Generalization is a function that joins different classes of objects with common characteristics to a higher level (generalized) class.
    - Measurement functions allow the calculation of distances, lengths, or areas.
2. **Overlay functions.** These belong to the most frequently used functions in a GIS application. They allow the combination of two (or more) spatial data layers comparing them position by position, and treating areas of overlap—and of non-overlap—in distinct ways. Many GISs support overlays through an algebraic language, expressing an overlay function as a formula in which the data layers are the arguments. In this way, we can find
3. **Neighbourhood functions.** Whereas overlays combine features at the same location, neighbourhood functions evaluate the characteristics of an area *surrounding* a feature's location. A neighbourhood function 'scans' the neighbourhood of the given feature(s), and performs a computation on it.
    - *Search functions* allow the retrieval of features that fall within a given *search window*. This window may be a rectangle, circle, or polygon.
    - *Buffer zone generation* (or buffering) is one of the best known neighbourhood functions. It determines a spatial envelope (*buffer*) around (a) given feature(s). The created buffer may have a fixed width, or a variable width that depends on characteristics of the area.
    - *Interpolation functions* predict unknown values using the known values at nearby locations. This typically occurs for continuous fields, like elevation, when the data actually stored does not provide the direct answer for the location(s) of interest. Interpolation of continuous data was discussed in Section 5.4.2.

	• *Topographic functions* determine characteristics of an area by looking at the immediate neighbourhood as well. Typical examples are slope computations on digital terrain models (i.e. continuous spatial fields). The *slope* in a location is defined as the plane tangent to the topography in that location. Various computations can be performed, such as: – determination of *slope angle*, – determination of *slope aspect*, – determination of *slope length*, – determination of *contour lines*. These are lines that connect points with the same value (for elevation, depth, temperature, barometric pressure, water salinity etc). 4. **Connectivity functions.** These functions work on the basis of networks, including road networks, water courses in coastal zones, and communication lines in mobile telephony. These networks represent spatial linkages between features. Main functions of this type include: • *Contiguity functions* evaluate a characteristic of a set of connected spatial units. One can think of the search for a contiguous area of forest of certain size and shape in a satellite image. • *Network analytic functions* are used to compute over connected line features that make up a network. The network may consist of roads, public transport routes, high voltage lines or other forms of transportation infrastructure. Analysis of such networks may entail *shortest path computations* (in terms of distance or travel time) between two points in a network for routing purposes. Other forms are to find all points reachable within a given distance or duration from a start point for allocation purposes, or determination of the capacity of the network for transportation between an indicated source location and sink location. • *Visibility functions* also fit in this list as they are used to compute the points visible from a given location (viewshed modelling or viewshed mapping) using a digital terrain model.
b	**Write a note on automatic classification. (Reference book Page 373)** **Automatic classification** User-controlled classifications require a classification table or user interaction. GIS software can also perform automatic classification, in which a user only specifies the number of classes in the output data set. The system automatically determines the class break points. Two main techniques of determining break points are in use. 1. *Equal interval technique*: The minimum and maximum values v_{min} and v_{max} of the classification parameter are determined and the (constant) interval size for each category is calculated as $(v_{max} - v_{min})/n$, where n is the number of classes chosen by the user. This classification is useful in revealing the distribution patterns as it determines the number of features in each category. 2. *Equal frequency technique*: This technique is also known as *quantile classification*. The objective is to create categories with roughly equal numbers of features per category. The total number of features is determined first and by the required number of categories, the number of features per category is calculated. The class break points are then determined by counting off the features in order of classification parameter value.

1	1	1	2	8
4	4	5	4	9
4	3	3	2	10
4	5	6	8	8
4	2	1	1	1

(a) original raster

1	1	1	1	4
2	2	3	2	5
2	2	2	1	5
2	3	3	4	4
2	1	1	1	1

(b) equal interval classification

original value	new value	# cells
1,2	1	9
3,4	2	8
5,6	3	3
7,8	4	3
9,10	5	2

1	1	1	2	5
3	3	4	3	5
3	2	2	2	5
3	4	4	5	5
3	2	1	1	1

(c) equal frequency classification

original value	new value	# cells
1	1	6
2,3	2	5
4	3	6
5,6	4	3
8,9,10	5	5

Figure 6.11: Example of two automatic classification techniques: (a) the original raster with cell values; (b) classification based on equal intervals; (c) classification based on equal frequencies. Below, the respective classification tables, with a tally of the number of cells involved.

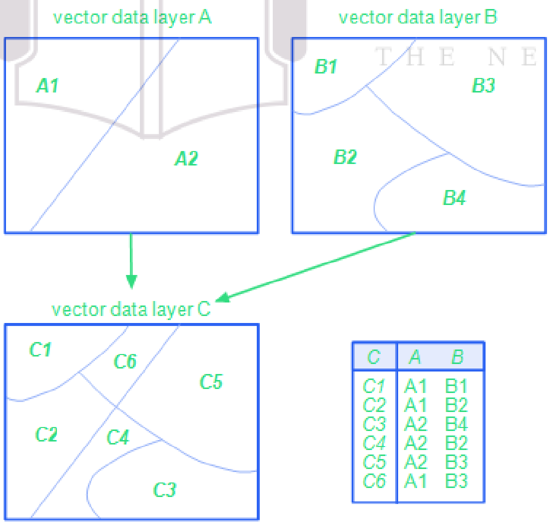
c **Explain vector overlay operations using suitable diagram. (Reference book Page 377)**

In the vector domain, overlay is computationally more demanding than in the raster domain. Here we will only discuss overlays from polygon data layers, but we note that most of the ideas also apply to overlay operations with point or line data layers.

The standard overlay operator for two layers of polygons is the *polygon intersection* operator. It is fundamental, as many other overlay operators proposed in the literature or implemented in systems can be defined in terms of it. The principles are illustrated in Figure.

The result of this operator is the collection of all possible polygon intersections; the attribute table result is a join—in the relational database the two input attribute tables. This output attribute table only contains one tuple for each intersection polygon found, and this explains why we call this operator a *spatial join*.

E-next THE NEXT LEVEL OF EDUCATION



C	A	B
C1	A1	B1
C2	A1	B2
C3	A2	B4
C4	A2	B2
C5	A2	B3
C6	A1	B3

Figure 6.12: The polygon intersect (overlay) operator. Two polygon layers A and B produce a new polygon layer (with associated attribute table) that contains all intersections of polygons from A and B. Figure after [8].

Two more polygon overlay operators are illustrated in Figure 6.14. The first is known as the *polygon clipping* operator. . It takes a polygon data layer and restricts its spatial extent to the generalized outer boundary obtained from all (selected) polygons in a second input layer. Besides this generalized outer boundary, no other polygon boundaries from the second layer play a role in the result.

A second overlay operator is *polygon overwrite*. The result of this binary operator is defined is a polygon layer with the polygons of the first layer, except where polygons existed in the second layer, as these take priority.

The fundamental operator of all these is *polygon intersection*. The others can be defined in terms of it, usually in combination with polygon selection and/or classification. For instance, the polygon overwrite of *A* by *B* can be defined as polygon intersection between *A* and *B*, followed by a (well-chosen) classification that prioritizes polygons in *B*, followed by a merge. The reader is asked to verify

this.
Vector overlays are usually also defined for point or line data layers. Their definition parallels the definitions of operators discussed above. Different GISs use different names for these operators, and one is advised to carefully check the documentation before applying any of these operators.

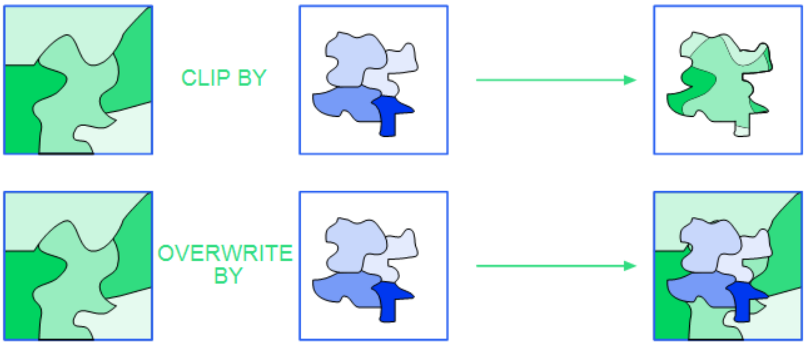


Figure 6.14: Two more polygon overlay operators: (a) polygon clip overlay clips down the left hand polygon layer to the generalized spatial extent of the right hand polygon layer; (b) polygon overwrite overlay overwrites the left hand polygon layer with the polygons of the right hand layer.

d **Perform the raster overlay operation**

R3 = CON(R1=3 AND (R2 => 45 and R2 < = 60) , 1 , 0)

R3 - Soil Type Raster

0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0
0	0	0	0	0	1	1	0	0
0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	1	0	0
0	0	1	0	0	0	1	0	0
0	0	1	1	1	1	1	0	0

e **Explain using example how Raster overlay operation can be performed using decision table? (Reference book Page 390)**

E-next THE NEXT LEVEL OF EDUCATION

Conditional expressions are powerful tools in cases where multiple criteria must be taken into account.
We could produce the output raster of Figure 6.19 with a map algebra expression such as:

Suitability := *CON ((Landuse = "Forest" AND Geology = "Alluvial") OR (Landuse = "Grass" AND Geology = "Shale"),* "Suitable", "Unsuitable")

some GISs accommodate setting up a separate decision table that will guide the raster overlay process. This extra table carries domain expertise, and dictates which combinations of input raster cell values will produce which output raster cell value. This gives us a raster overlay operator using a decision table, as illustrated in Figure 6.19. The GIS will have supporting functions to generate the additional table from the input rasters, and to enter appropriate values in the table.

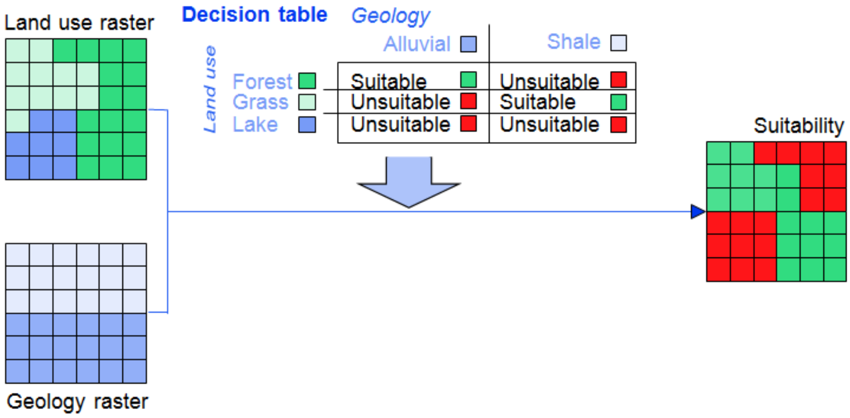


Figure 6.19: The use of a decision table in raster overlay. The overlay is computed in a suitability study, in which land use and geology are important factors. The meaning of values in both input rasters, as well as the output raster can be understood from the decision table.

f **List any five examples where advanced computations on continuous fields are required. (Reference book Page 406)**

	There are numerous examples where more advanced computations on continuous field representations are needed. A short list is provided below. • *Slope angle calculation* The calculation of the slope steepness, expressed as an angle in degrees or percentages, for any or all locations. • *Slope aspect calculation* The calculation of the aspect (or orientation) of the slope in degrees (between 0 and 360 degrees), for any or all locations. • *Slope convexity/concavity calculation* Slope convexity—defined as the change of the slope (negative when the slope is concave and positive when the slope is convex)—can be derived as the second derivative of the field. • *Slope length calculation* With the use of neighbourhood operations, it is possible to calculate for each cell the nearest distance to a watershed boundary (the upslope length) and to the nearest stream (the downslope length). This information is useful for hydrological modelling. • *Hillshading* is used to portray relief difference and terrain morphology in hilly and mountainous areas. The application of a special filter to a DEM produces hillshading. • *Three-dimensional map display* With GIS software, three-dimensional views of a DEM can be constructed, in which the location of the viewer, the angle under which s/he is looking, the zoom angle, and the amplification factor of relief exaggeration can be specified. Three-dimensional views can be constructed using only a predefined mesh, covering the surface, or using other rasters (e.g. a hillshading raster) or images (e.g. satellite images) which are draped over the DEM. • *Determination of change in elevation through time* The cut-and-fill volume of soil to be removed or to be brought in to make a site ready for construction can be computed by overlaying the DEM of the site before the work begins with the DEM of the expected modified topography. It is also possible to determine landslide effects by comparing DEMs of before and after the landslide event. • *Automatic catchment delineation* Catchment boundaries or drainage lines can be automatically generated from a good quality DEM with the use of neighbourhood functions. The system will determine the lowest point in the DEM, which is considered the outlet of the catchment. From there, it will repeatedly search the neighbouring pixels with the highest altitude. This process is continued until the highest location (i.e. cell with highest value) is found, and the path followed determines the catchment boundary. For delineating the drainage network, the process is reversed. Now, the system will work from the watershed downwards, each time looking for the lowest neighbouring cells, which determines the direction of water flow. • *Dynamic modelling* Apart from the applications mentioned above, DEMs are increasingly used in GIS-based dynamic modelling, such as the computation of surface run-off and erosion, groundwater flow, the delineation of areas affected by pollution, the computation of areas that will be covered by processes such as debris flows and lava flows. • *Visibility analysis* A viewshed is the area that can be 'seen'—i.e. is in the direct line-of-sight—from a specified target location. Visibility analysis determines the area visible from a scenic lookout, the area that can be reached by a radar antenna, or assesses how effectively a road or quarry will be hidden from view.
5.	**Attempt *__any three__* of the following:**
a.	**What is the relationship between Map and GIS? (Reference book Page 441)** There is a strong relationship between maps and GIS. More specifically, maps can be used as input for a GIS. • Where • What

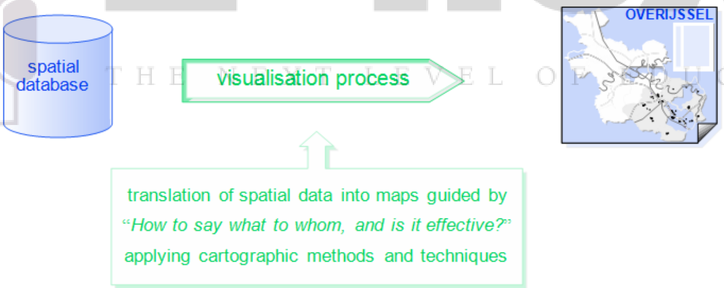
- When

As soon as a question contains a “where?” question, a map can often be the most suitable tool to solve the question and provide the answer. “Where do I find Enschede?” and “Where did ITC’s students come from?” are both examples. Of course, the answers could be in non-map form like “in the Netherlands” or “from all over the world.” These answers could be satisfying, however, they do not give the full picture.
A map would put these answers in a spatial context. It could show where in the Netherlands Enschede is to be found and where it is located with respect to Schiphol–Amsterdam airport, where most students arrive. A world map would refine the answer “from all over the world,” since it reveals that most students arrive from Africa and Asia, and only a few come from the Americas, Australia and Europe.
As soon as the location of geographic objects (“where?”) is involved, a map becomes useful. However, maps can do more than just providing information on location. They can also inform about the thematic attributes of the geographic objects located in the map. An example would be “What is the predominant land use in southeast Twente?” The answer could, again, just be verbal and state “Urban.” However, such an answer does not reveal patterns. Maps can answer the “What?” question only in relation to location.
A third type of question that can be answered from maps is related to “When?” For instance, “When did the Netherlands have its longest coastline?” The answer might be “1600,” and this will probably be satisfactory to most people. However, it might be interesting to see how this changed over the years. A set of maps could provide the answer

b. **Explain the visualization process in GIS. (Reference book Page 152)**

The cartographic visualization process is used for translation or conversion of spatial data from a database into graphics.

E-next THE NEXT LEVEL OF EDUCATION



The producer of these visual products may be a professional cartographer, but may also be a discipline expert, for instance, mapping vegetation stands using remote sensing images, or health statistics in the slums of a city. To enable the translation from spatial data into graphics, we assume that the data are available and that the spatial database is well-structured.
The visualization process can vary greatly depending on where in the spatial data handling process it takes place and the purpose for which it is needed. Visualizations can be, and are, created during any phase of the spatial data handling process as indicated before. They can be simple or complex, while the production time can be short or long.
The visualization process is guided by the question “How do I say what to whom?” “How” refers to cartographic methods and techniques. “I” represents the cartographer or map maker, “say” deals with communicating in graphics the semantics of the spatial data. “What” refers to the spatial data and its characteristics, (for instance, whether they are of a qualitative or quantitative nature). “Whom” refers to the map audience and the purpose of the map—a map for scientists requires a different approach than a map on the same topic aimed at children.

c. **What are Bertin’s six categories of visual variables? (Reference book Page 466)**

Bertin distinguished six categories, which he called the *visual variables* and which may be applied to point, line and area symbols.

- *Size*,
- *Value (lightness)*,
- *Texture*,
- *Colour*,
- *Orientation* and
- *Shape*.

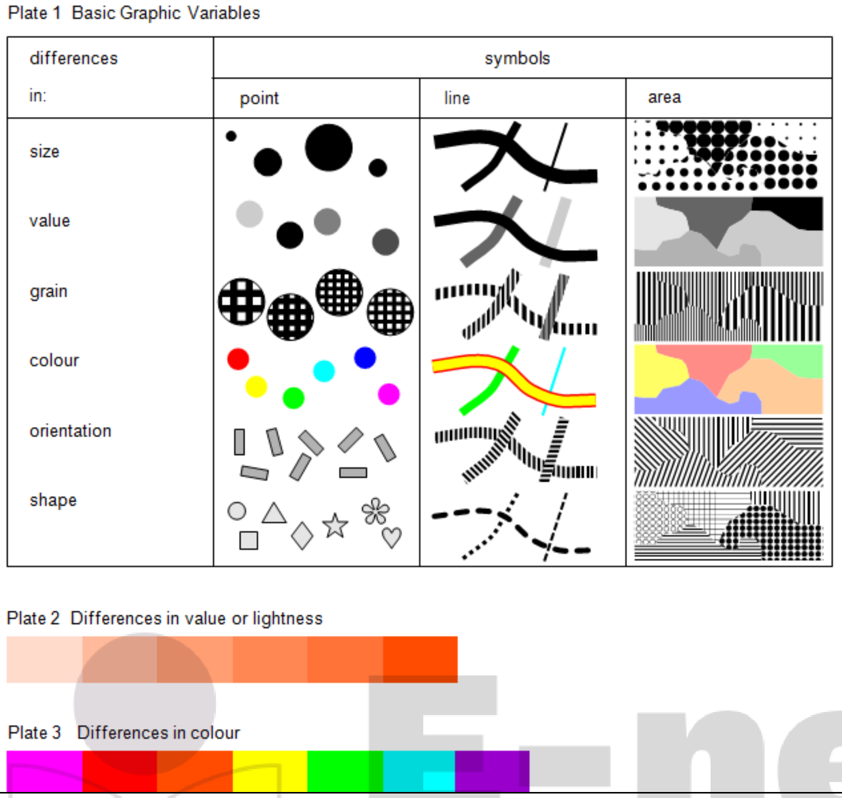


Figure 7.11: Bertin's six visual variables illustrated. Source: Plate 1 in [31].

E-next
THE NEXT LEVEL OF EDUCATION

d. **How to map terrain elevation? Explain. (Reference book Page 477)**

Terrain elevation can be mapped using different methods. Often, one will have collected an elevation data set for individual points like peaks, or other characteristic points in the terrain. Obviously, one can map the individual points and add the height information as text. However, a *contour map*, in which the lines connect points of equal elevation, is generally used. To visually improve the information content of such a map the space between the contour lines can be filled with colour and value information following a convention, e.g. green for low elevation and brown for high elevation areas. This technique is known as hypsometric or layer tinting. Even more advanced is the addition of *shaded relief*.

Interactive functions are required to manipulate the map in three-dimensional space in order to look behind some objects. These manipulations include panning, zooming, rotating and scaling. Scaling is needed, particularly along the *z*-axis, since some maps require small-scale elevation resolution, while others require large-scale resolution, i.e. vertical exaggeration.

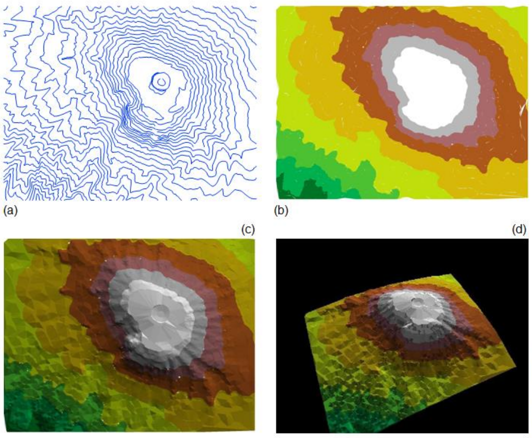


Figure 7.18: visualization of terrain elevation: (a) contour map; (b) map with layer tints; (c) shaded relief map; (d) 3D view of the terrain

e.	**How to distinguish between three temporal cartographic techniques? Explain. (Reference book Page 43)** It is possible to distinguish between three temporal cartographic techniques 1. *Single static map:* Specific graphic variables and symbols are used to indicate change or represent an event. Figure 7.20(a) applies the visual variable value to represent the age of the built-up areas; 2. *Series of static maps:* A single map in the series represents a 'snapshot' in time. Together, the maps depict a process of change. Change is perceived by the succession of individual maps depicting the situation in successive snapshots. It could be said that the temporal sequence is represented by a spatial sequence, which the user has to follow, to perceive the temporal variation. The number of images should be limited since it is difficult for the human eye to follow long series of maps (Figure 7.20(b)); 3. *Animated map:* Change is perceived to happen in a single image by displaying several snapshots after each other just like a video cut with successive frames. The difference with the series of maps is that the variation can be deduced from real 'change' in the image itself, not from a spatial sequence For the user of a cartographic animation, it is important to have tools available that allow for interaction while viewing the animation. Seeing the animation play will often leave users with many questions about what they have seen. Just replaying the animation is not sufficient to answer questions like "What was the position of the coastline in the north during the 15th century?"
f.	**Write a note on Map Cosmetics. (Reference book Page 485)** Most maps in this chapter are correct from a cartographic grammar perspective. However, many of them lack the additional information needed to be fully understood that is usually placed in the margin of printed maps. Each map should have, next to the map image, a *title*, informing the user about the topic visualized. A *legend* is necessary to understand how the topic is depicted. Additional marginal information to be found on a map is a *scale indicator*, a *north arrow* for orientation, the *map datum* and *map projection* used, and some *lineage* information, (such as data sources, dates of data collection, methods used, etc.). Further information can be added that indicates when the map was issued, and by whom (author / publisher). All this information allows the user to obtain an impression of the quality of the map, and is comparable with metadata describing the contents of a database or data layer. On paper maps, these elements (if all relevant) have to appear next to the map face itself. Maps presented on screen often go without marginal information, partly because of space constraints. However, on-screen maps are often interactive, and clicking on a map element may reveal additional information from the database. Legends and titles are often available on demand as well. Maps constructed via the basic cartographic guidelines are not necessarily visually appealing maps. Although well-constructed, they might still look sterile. The design aspect of creating appealing maps also has to be included in the visualization process. 'Appealing' does not only mean having nice colours. One of the keywords here is *contrast*. Contrast will increase the communicative role of the map since it creates a hierarchy in the map contents, assuming that not all information has equal importance. This design trick is known as *visual hierarchy* or the figure-ground concept. The need for visual hierarchy in a map is best understood. The map of the ITC building and surroundings in part (b) is an example of a map that has visual hierarchy applied. The first object to be noted will be the ITC building (the darkest patches in the map) followed by other buildings, with the road on a lower level and the parcels at the lowest level.
	______________END_________________

E-next